**Project Proposal: Heart Disease Prediction**

Heart disease is a leading cause of death worldwide. An early and accurate prediction can significantly improve patient outcomes by enabling preventive measures and timely interventions. This project aims to develop a machine-learning model to predict the likelihood of heart disease in patients using clinical and demographic data.

In this project, I retrieve a dataset encapsulating various health metrics from heart patients, including age, blood pressure, heart rate, and more. The objective is to develop a predictive model to accurately identify individuals with heart disease. The primary emphasis is on ensuring that the model identifies all potential patients, making recall for the positive class a crucial metric.

My key objectives are the following:

* Develop an accurate prediction model using machine learning algorithms.
* Identify critical factors that contribute to the risk of heart disease.
* Provide a user-friendly interface for healthcare professionals to use the prediction tool.

I will access the datasets by logging in to KaggeI and exploring them, uncovering patterns, distributions, and relationships within them. Then, I will conduct extensive exploratory data analysis (EDA), Diving deep into bivariate relationships against the target. I will follow the following steps:

* Remove irrelevant features
* Address missing values
* Treat outliers
* Encode categorical variables
* Transform skewed features to achieve normal-like distributions

Establish the pipelines for models that require scaling to build. Implement and tune classification models, including KNN, SVM, Decision Trees, and Random Forests. Emphasize achieving high recall for class 1 to ensure comprehensive identification of heart patients. Evaluate and Compare Model Performance, such as utilizing precision, recall, and F1-score to gauge models' effectiveness.

The following methodology will be followed to evaluate the datasets:

I will use publicly available datasets in the Kaggle, such as the Heart Disease Prediction dataset. I will utilize the steps such as handling missing values, normalizing data, and encoding categorical variables to evaluate the data.

Various machine learning algorithms, such as Logistic Regression, Decision Trees, Random Forests, and Neural Networks, will be explored. The models will be trained (supervised) and validated using techniques such as cross-validation. Model performance will be assessed using accuracy, precision, recall, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

I will utilize the following tools and technologies:

* Programming Languages: Python
* Machine Learning Frameworks: Scikit-learn, TensorFlow, Keras
* Data Handling: Pandas, NumPy

The expected outcomes are the following:

* A reliable model for heart disease prediction.
* Insights into significant risk factors for heart disease.

The budget will cover data access, computational resources, and personnel expenses. Upon further detailing, an estimated budget breakdown will be provided.

**Conclusion**

This project aims to leverage machine learning to improve heart disease prediction, thereby contributing to better healthcare outcomes. We anticipate that the developed model will aid healthcare professionals in making informed decisions, ultimately saving lives.